

Income Estimation for Former Household Members in the Redesigned SIPP

A Methodological Analysis of Instrument Efficacy

By Lindsay Monte & Ashley Edwards | Social, Economic, and Housing Statistics Division | U.S. Census Bureau, U.S. Department of Commerce

Background

The Census Bureau's Survey of Income and Program Participation (SIPP) is a longitudinal survey, collecting monthly data on family composition and economic well-being. In 2006, Congress mandated a reengineering of the SIPP (NRC, 2009). As part of this redesign, the interview schedule moved from every four months to a single annual interview. The redesigned survey utilizes an Event History Calendar as a memory aid, and is known as the Survey of Income and Program Participation – Event History Calendar, or the SIPP-EHC.

One of the challenges of the SIPP-EHC is capturing data on individuals who were part of the household at some point in the reference year, but left the household before the interview. Data on these individuals, referred to as "Type2" people, are important in order to understand how household composition and income change over the year.

Research Questions

- How well does the SIPP-EHC capture Type2 people when compared to SIPP data?
 - Are the Type2 people captured in the SIPP-EHC demographically and economically similar to similar individuals captured in the SIPP?
- Are we accurately measuring annual income amounts for Type2 people in the SIPP-EHC?
 - Are there ways that we could improve our battery of Type2 questions in the SIPP-EHC to improve income imputation, and better understand household financial well-being?

Data

We compare data on Type2 persons collected from the 2010 SIPP-EHC field test (n=714) to data collected through the 2008 SIPP panel.

Our comparison sample in the 2008 SIPP consists of individuals who would have been considered Type2 persons under the SIPP-EHC survey design, referred to as "Pseudo-Type2" persons (n=455).

Both samples consist of individuals who do not live at the sampled address at the time of the Q1 2010 interview, but who, at some point between January and December 2009, lived with someone who resides at the sampled address as of the Q1 2010 interview.¹

Methods

Bayesian Bootstrap Multiple Imputation model to assess income

- Two imputation models:
 - Using data available from the Type2 question battery
 - Using question battery data, plus additional demographics²
- Four implicates of each model
- Results are the means across these implicates

Conclusions/Implications

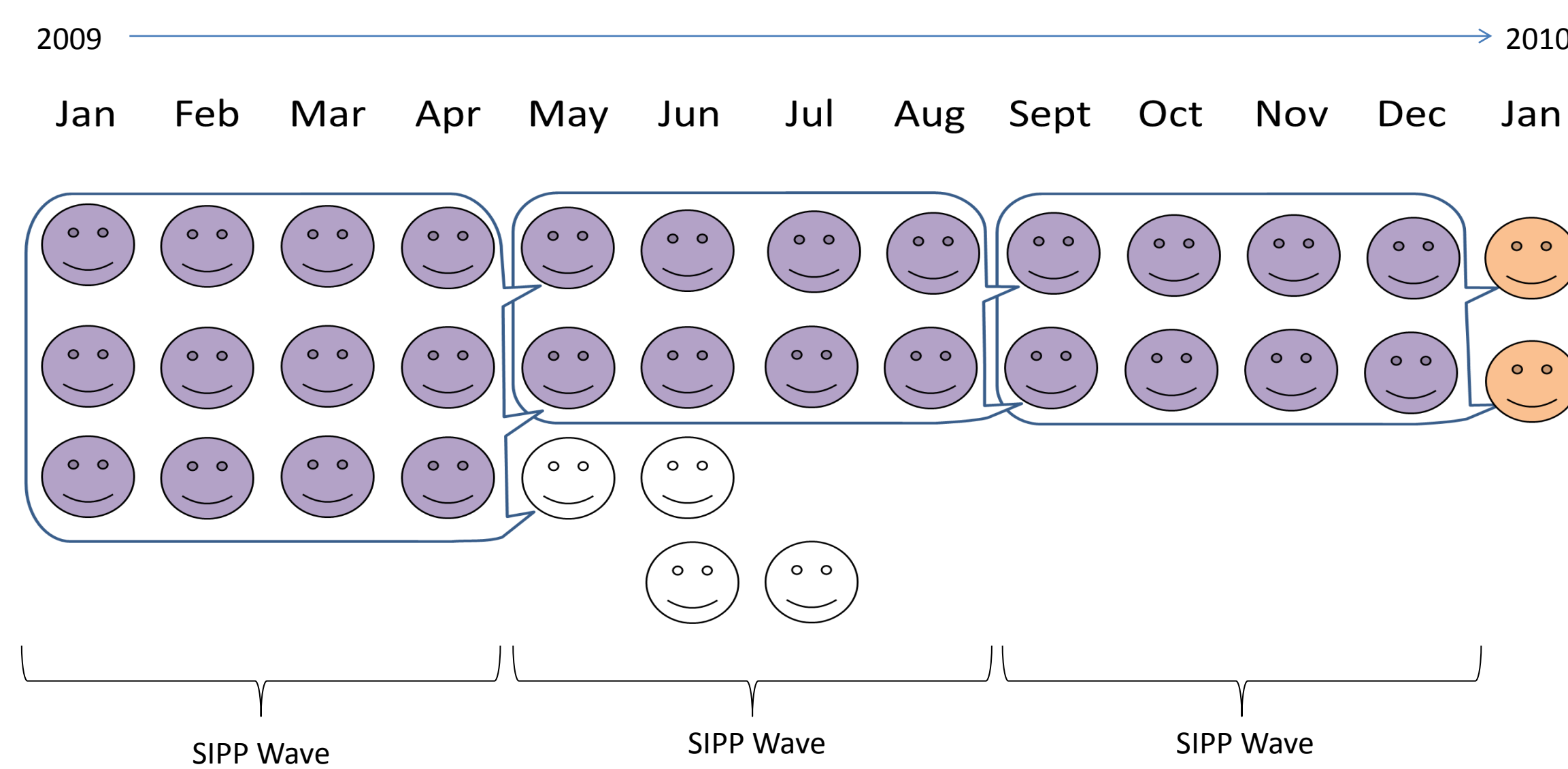
The SIPP-EHC captures fewer Type2s overall than does the SIPP (Fig.1), although the SIPP-EHC captures more short stay Type2s (Fig.9). Type2s are much more likely to be non-relatives of respondents than are Pseudo-Type2s (Fig.4), and are also distinct in a number of other ways (Figs.2,3,5,8), suggesting that the SIPP-EHC is not necessarily capturing the same individuals as the SIPP.

Despite these differences, we do not find significant differences in mean reported income between samples (Fig. 11), although Type2s are more likely to be reported as having annual incomes greater than \$50,000. (Fig.6). However, for approximately 40% of the SIPP-EHC Type2s, we did not obtain a useable income response (Fig.12). Moreover, we find some significant demographic differences between Type2s for whom we have reports of income and those for whom we do not (Fig.13).

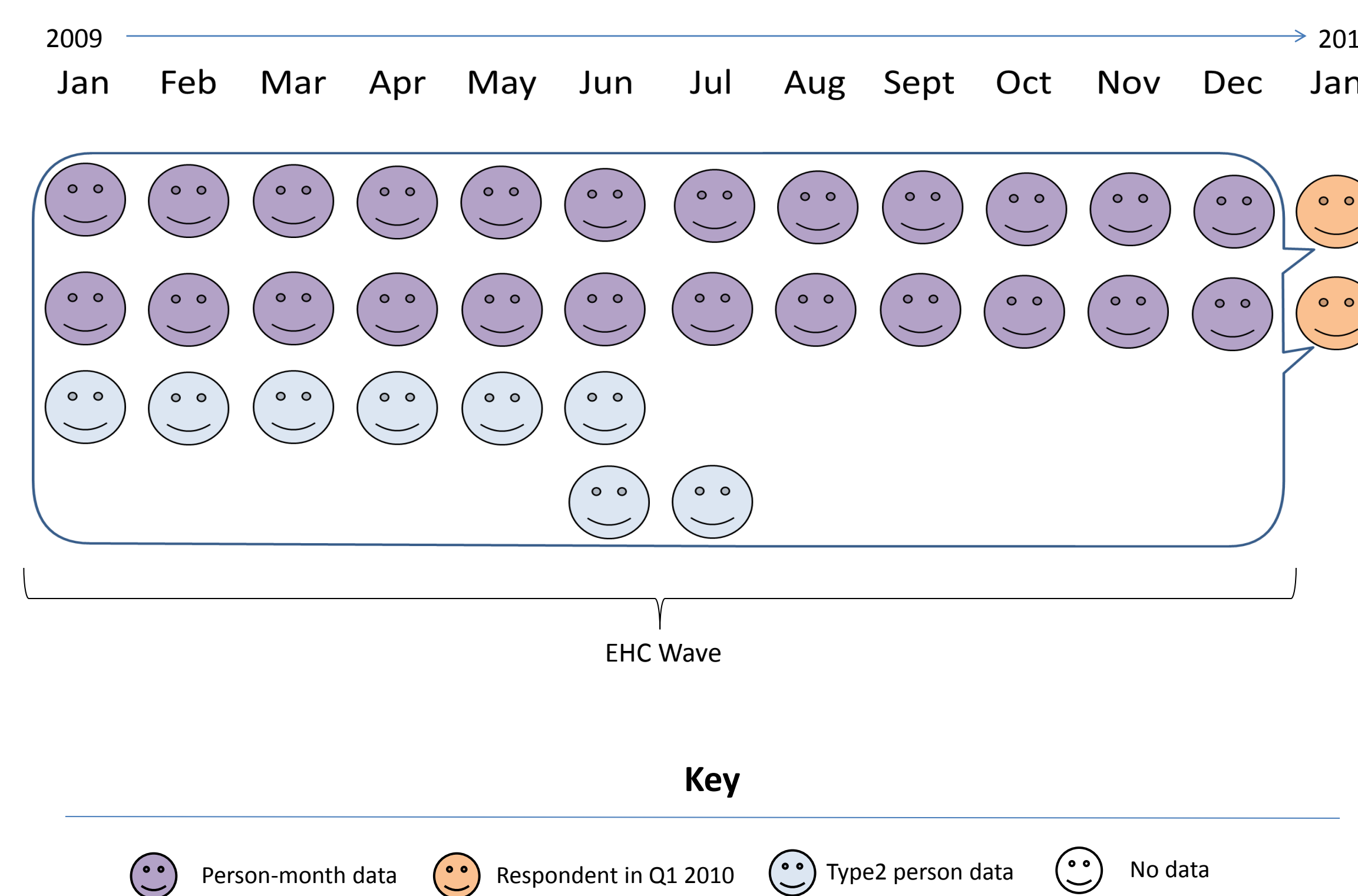
Nonetheless, regardless of how we consider non-response in our imputations, we do not find any significant differences in mean imputed income, either when compared to the SIPP mean or the mean of SIPP-EHC Type2 reports (see Model 1, Table 2). We further find that the addition of derived demographic information does not change our income imputations (see Model 2, Table 2).

This suggests that, for the Type2s captured, the current question battery may be sufficient, and follow-up income questions may not be necessary (Fig.10). However, additional efforts may be needed to capture Type2s whose co-residence was early in the year (Fig.7), and additional research is needed to understand differential non-response by demographic categories (Fig.13).

Pseudo-Type2s in the SIPP

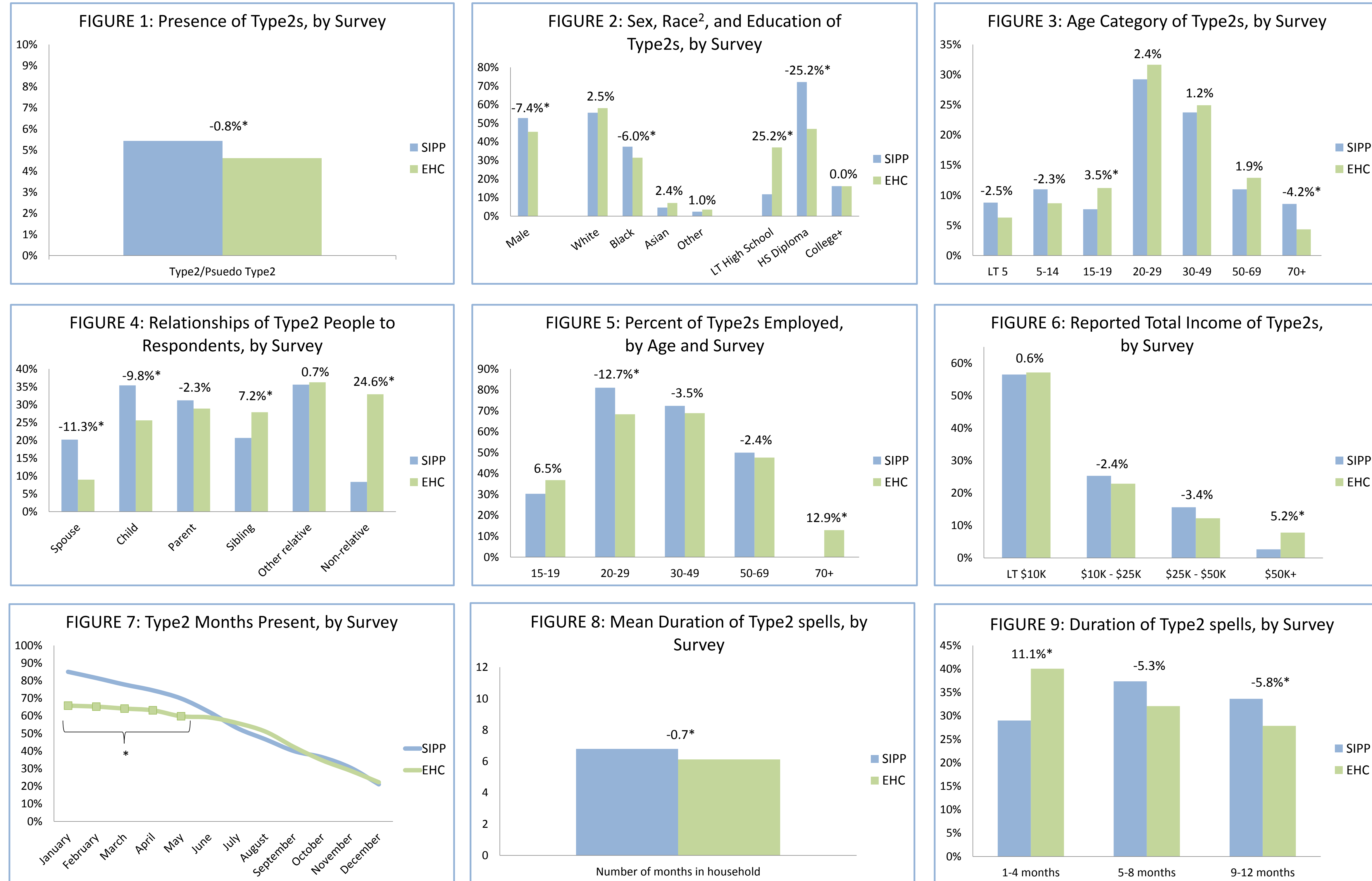


Type2s in the SIPP-EHC



Key: Person-month data, Respondent in Q1 2010, Type2 person data, No data

Who are our "Type2" people and how do they compare across surveys?



Economic Analysis

FIGURE 10: Income Questions from the SIPP-EHC Type2 Question Battery

Asked about Type2 individuals age 15 and older:
Including all income sources, what would you guess was (Type2 person's) annual income?
 > IF VALID DOLLAR AMOUNT: Reported income
 > IF DON'T KNOW / REFUSE:
 - Was it more than \$25,000?
 - Was it more than \$50,000?
 - Was it less than \$10,000?
 IF DON'T KNOW / REFUSE: Non-response

FIGURE 11: Mean Type2 Reported Income by Survey



FIGURE 12: Reporting of Total Income of Type2s in SIPP-EHC

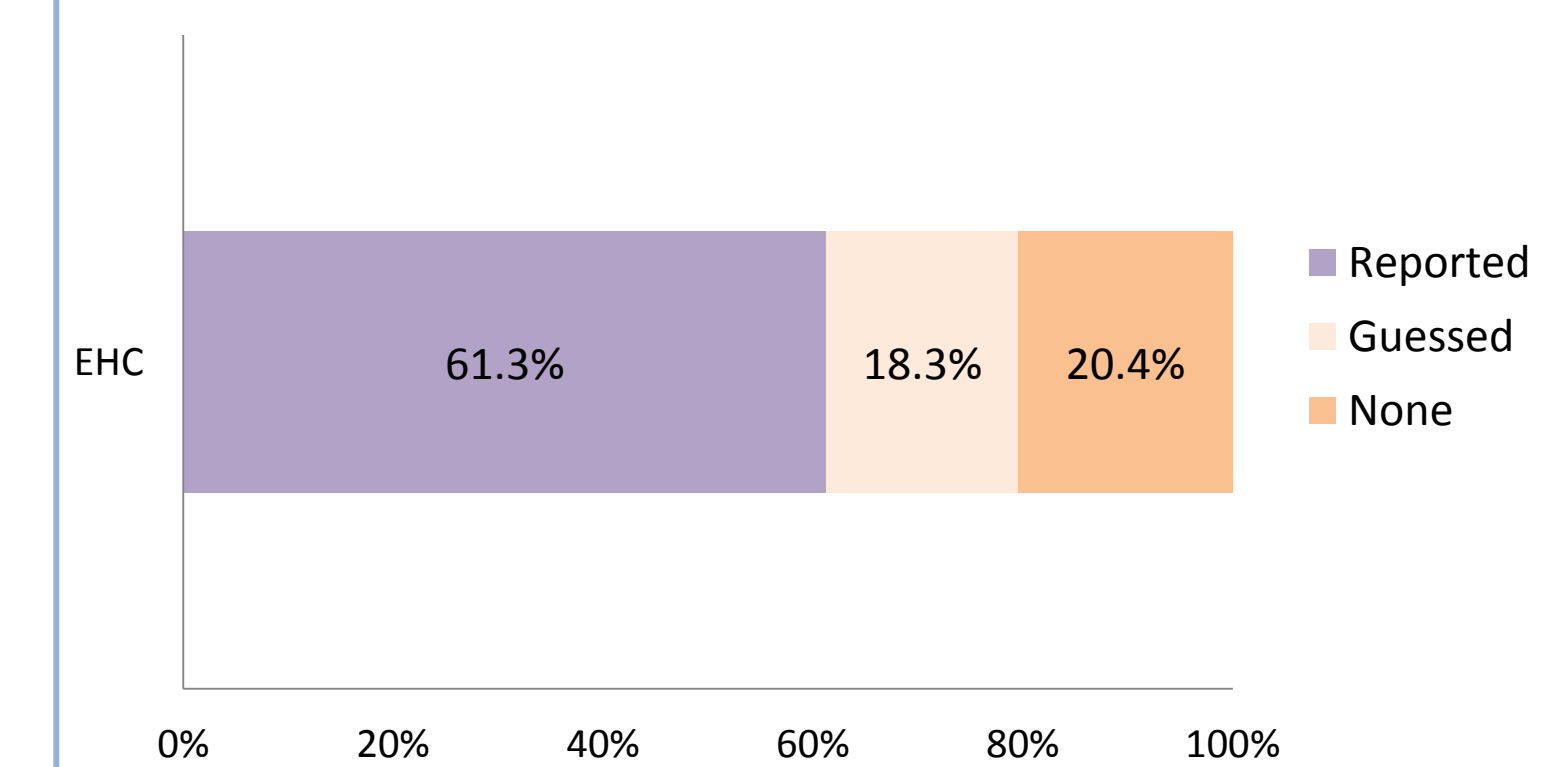
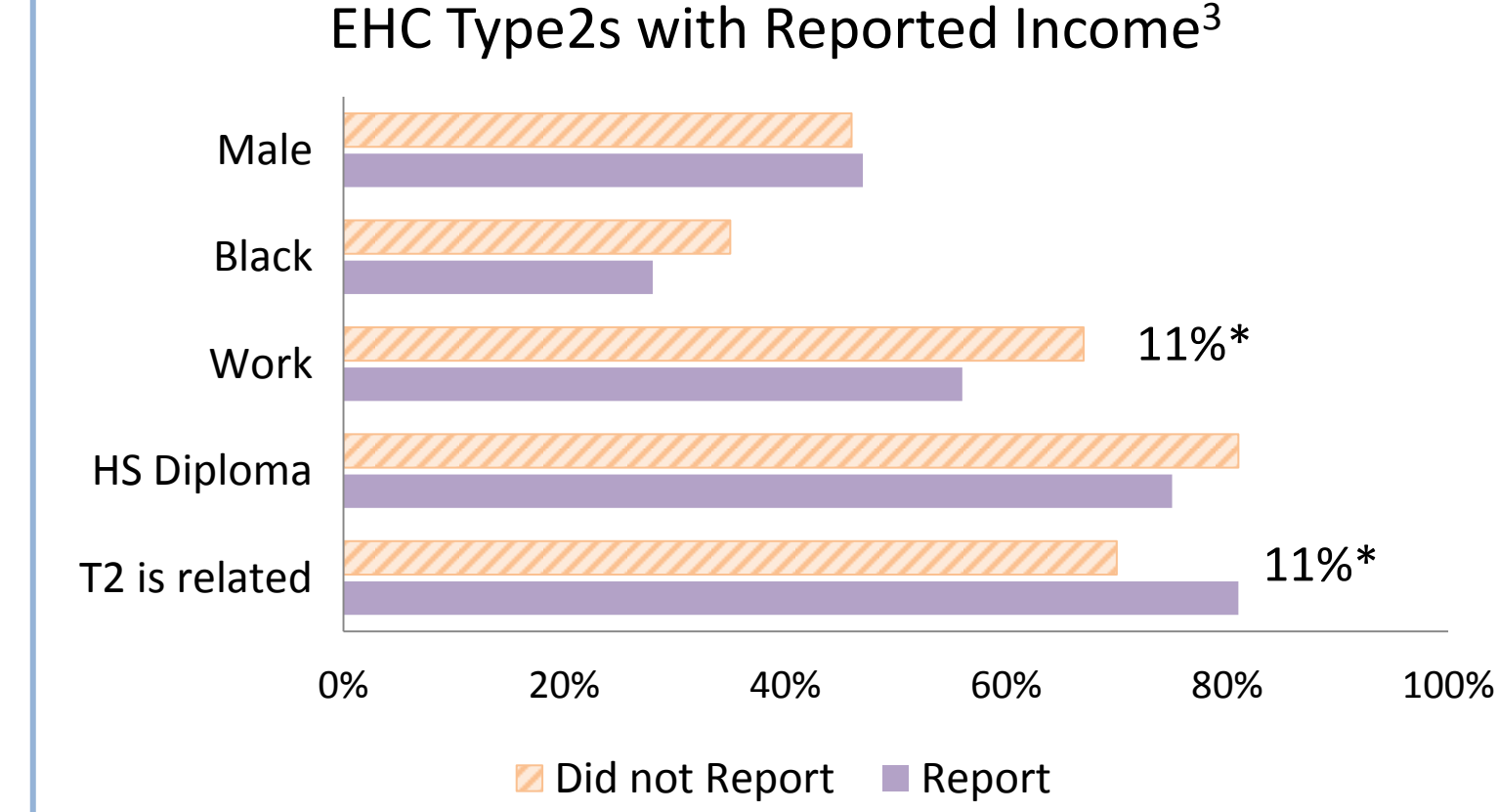


FIGURE 13: Demographic Differences for SIPP-EHC Type2s with Reported Income³



* Defined as statistically significant at the 95 percent confidence level
 DATA SOURCE: 2008 SIPP Panel and 2010 SIPP-EHC. <http://www.census.gov/sipp/source.html>
 NOTES:
 1. These data collected from the SIPP 2008 panel have been subset to match the 2009 calendar year, sampling characteristics, and geography of the 2010 SIPP-EHC field test. For this reason, the data presented here are not intended to be nationally representative. Differences in samples composition as well as the opportunity for attrition in the 2008 SIPP panel may introduce selection bias into our analysis.
 2. Because race and marital status are not part of the Type2 questions, we logically impute them from available household data.
 3. Age and marital status were also compared but the differences are not significant.
 WORKS CITED:
 National Research Council. Reengineering the Survey of Income and Program Participation. 2009. http://www.nap.edu/catalog.php?record_id=12715

Model Results

Table 1. Imputation Model Design

Independent Variables	SIPP	EHC	Model 1	Model 2	Outcome Variable	
					Variable	Definition
Sex	Available	Available	X	X	Annual person income from all sources	Defined as the sum of personal earned income, transfer income, and other income
Education	Available	Available	X	X		
Age	Available	Available	X	X		
Employment/Work	Available	Available	X	X		
Marital Status ²	Available	Logically derived		X		
Race ²	Available	Logically derived		X		

Table 2. Imputation Model Results

Model	Imputed	EHC	EHC Dif.	95% Sig.	SIPP		SIPP Dif.	95% Sig.
					Mean	Std. Err.		
Model 1	All Type2 imputed	16,153	780	2,463	ns		-459	ns
	Only missing Type2 imputed	17,738	1,269	878	ns		-2,043	ns
	Only non-missing Type2 imputed	15,152	977	3,464	ns	15,694	860	543
	Missing T2 imputed, non-missing T2 retained	18,276	1,131	340	ns		-2,581	ns
Model 2	All Type2 imputed	14,662	656	3,954	ns		1,032	ns
	Only missing Type2 imputed	15,085	1,069	3,531	ns		609	ns
	Only non-missing Type2 imputed	14,395	831	4,221	ns	15,694	860	1,300
	Missing T2 imputed, non-missing T2 retained	17,249	1,099	1,367	ns		-1,555	ns

The views expressed in this research, including those related to statistical, methodological, technical, or operational issues, are solely those of the authors and do not necessarily reflect the official positions or policies of the Census Bureau. The authors accept responsibility for all errors. This poster is released to inform interested parties of ongoing research and to encourage discussion of work in progress.
 The estimates in this poster are based on responses from a sample of the population and may differ from actual values because of sampling variability or other factors.